

UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

In re. PHARMACEUTICAL INDUSTRY)	MDL No. 1456
AVERAGE WHOLESALE PRICE)	Civil Action No. 01-12257-PBS
LITIGATION)	
)	Hon. Patti Saris
THIS DOCUMENT RELATES TO THE)	
AMENDED MASTER CONSOLIDATED)	
CLASS ACTION)	
)	

**REPORT OF UNITED STATES DEPARTMENT OF
HEALTH AND HUMAN SERVICES CONCERNING
SUBPOENA FOR MEDICARE BENEFICIARY DATA**

The United States Department of Health and Human Services (DHHS), through its undersigned counsel, respectfully submits this report to the Court concerning production of data requested by the subpoena duces tecum dated June 19, 2006, served on the DHHS, Office of General Counsel, on June 7, 2006, and concerning the Court's Order of June 20, 2006, in which the Court asked the government to provide certain information about the production of data.

At the hearing on June 26, 2006, the undersigned counsel George Henderson, appearing for the DHHS, advised the Court that it would take about three months for the DHHS to provide beneficiary information for the period 1998 to 2005. Counsel also indicated that it would take essentially the same amount of time to provide beneficiary

information for the longer period 1991 to 2005. Subsequent to the hearing the undersigned counsel learned that the latter estimate was incorrect.¹

In an effort to provide the Court and the parties with more information, two conference calls have been conducted between officials of the DHHS, the U.S. Attorney's Office, and various counsel for the plaintiffs and defendants. The second conference call, on June 30, 2006, included DHHS employee John Fletcher, whose declaration is attached hereto, the U.S. Attorney's Office, counsel for the plaintiffs (Edward Notargiacomo), representatives of Complete Claims Solutions (a vendor who would issue class notice), and counsel for the defendants (James Zucker (Bristol-Myers Squibb), Michael Flynn (AstraZeneca)). A draft declaration was provided to participating counsel for the parties in advance of the conference call. Participating counsel for the parties were permitted to ask whatever questions they wished of Mr. Fletcher, and Mr. Fletcher answered to the best of his abilities. Mr. Fletcher explained the format in which the data would be produced, and it was agreed that DHHS would provide the data to Complete Claims Solutions in a text format that would facilitate its use for making mailing labels.

Through this report to the Court, the DHHS is attempting to respond to all questions to the extent practical. The DHHS believes its response is complete and that there is no need for Mr. Fletcher to be diverted from his official responsibilities to travel to Boston to testify on such matters. Counsel for the plaintiffs and counsel for the

¹ Counsel sincerely regrets the error. It was due to a mis-communication with the client agency.

defendants who participated in the June 30 conference call declined to state whether they would have additional questions of Mr. Fletcher at a hearing. The DHHS respectfully requests that the Court advise the DHHS soon whether Mr. Fletcher needs to be present to testify at the hearing, so that travel plans can be made.

Attached hereto as Exhibit A is the Declaration of John A. Fletcher (Fletcher Dec.), the Director of the Division of Medicare and Medicaid Utilization Data Development (DMUD) within the Office of Information Services for the Centers for Medicare & Medicaid Services (CMS). In his declaration Mr. Fletcher describes in detail the computer processing facilities used by CMS for its operations, the procedures followed for processing data requests, the estimated times needed to produce data under several different scenarios, and why the times cannot be shortened or made more definite.

As explained by Mr. Fletcher, the simplest production of data would consist of names and addresses of beneficiaries on whose behalf a claim was filed with Medicare for any of the drugs identified in the subpoena (approximately 47 HCPCS codes²), covering the period 1998 to January 1, 2005. Fletcher Dec. ¶ 17.a. Production of this data would take three months. Beneficiaries would not be matched with particular drugs. If the time period were expanded to include 1991 - 1998 (15 years of data), the amount of time

² The Healthcare Common Procedure Coding System (HCPCS) is a format in which drugs and biologicals may be coded for reimbursement and other reporting purposes.

would double, to six months. The cost of producing 15 years worth of information would be \$109,050, plus \$3,075 for each increment of three million names and addresses. Id.

If the data request were modified to require a matching of beneficiaries to the relevant drugs (of which there are 27), this would increase the time needed to produce the information to eight months, for all 15 years. Fletcher Dec. ¶ 17.b. If only 1998 - 2005 data were needed, it would take five months. Id. The estimated cost would be \$2,944,350, plus \$3,075 for each increment of three million names and addresses. If the data extraction were limited to the time period 1998 - 2005, it would take five months. Id.

If the data request were further modified to allow a matching of beneficiaries to particular HCPCS codes (there are 47 codes), it would take approximately 10 months to produce the data for the 15 years 1991 - 2005. The cost of producing this data would be \$5,125,350, plus \$3,075 for each increment of three million names and addresses. Id. ¶ 17.c. If the data is extracted only for the period 1998 - 2005, it would take approximately six months. Id.

Although the times estimated for production appears to be longer than the Court would like, it is not feasible to produce the data sooner. All of CMS' processing and data systems are run on one mainframe computer. Fletcher Dec. ¶ 3. The primary purpose of this computer is to support the information technology infrastructure required to meet the mission of the DHHS. Numerous critical systems rely on the smooth operation of this computer, including Medicare enrollment, Social Security information exchanges, drug

data pricing, Part D (the new drug benefit), Medicare Advantage, and Medicare claims and payments. Id. The mainframe computer's critical operations cannot be stopped or slowed without severe and unacceptable consequences. Id. ¶¶ 5-6. As a result, there is limited computer processing capacity available for other projects. Of the other projects that are run on the mainframe computer, nearly 90 percent are for important agency business, including tracking information for fraud and abuse purposes, tracking program utilization, and providing statistical information for CMS' use. This work is critical to the integrity of the Medicare program. Id. ¶ 6.

The DHHS recognizes that a prompt resolution of the claims on behalf of certain Medicare beneficiaries in this case is desirable. Accordingly, once the DHHS is advised of the specific parameters of the information being requested, the agency is willing to proceed promptly with processing the request.

Respectfully submitted,

MICHAEL J. SULLIVAN
United States Attorney

By: /s/ George B. Henderson, II
George B. Henderson, II
Damian W. Wilmot
Assistant U.S. Attorneys
John J. Moakley U.S. Courthouse
1 Courthouse Way, Suite 9200
Boston, MA 02110
(617) 748-3272, 3398

Exhibit A

UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

In re. PHARMACEUTICAL INDUSTRY)	MDL No. 1456
AVERAGE WHOLESALE PRICE)	Civil Action No. 01-12257-PBS
LITIGATION)	
)	Hon. Patti Saris
THIS DOCUMENT RELATES TO THE)	
AMENDED MASTER CONSOLIDATED)	
CLASS ACTION)	
)	

DECLARATION OF JOHN A. FLETCHER

I, John A. Fletcher, declare as follows:

1. I have held the position of Director, Division of Medicare and Medicaid Utilization Data Development (DMUD) within the Office of Information Services for the Centers for Medicare & Medicaid Services (CMS) since 2004. DMUD is the CMS component responsible for loading, storing and housing all CMS data. Prior to my current position, I was the Director of the Division of Data Access and Business Intelligence (DDABI) for one year. Most notably, I was also the Project Manager for the Data Extract System (DESY) for five years which encompassed the development and implementation of the system. The statements made in this declaration are based on my personal knowledge, information contained in agency files, and information furnished to me in the course of my official duties.

2. I understand that a subpoena has been issued seeking the names and addresses of Medicare beneficiaries on whose behalf Part B drug claims for various drugs have been submitted to Medicare for payment, covering the period 1991 to January 1, 2005. I understand that the Court and parties also seek an explanation regarding the cost and length of time required to produce this information. In the paragraphs below, I attempt to provide this explanation.

3. All of CMS' processing and data systems are run on one mainframe computer located at the CMS Central office in Baltimore, Maryland. The mainframe runs at 100% capacity 24 hours a day, 7 days a week. The mainframe is periodically shut down for upgrades, maintenance and mission critical projects such as uploading the Part D enrollment database. The mainframe is critical to the day to day administration of the Medicare program. Its primary purpose is to support the information technology infrastructure required to meet the agency mission. In total, there are 22 major databases (and approximately 330 additional databases) or systems which run concurrently on the mainframe -including critical database systems such as Medicare enrollment, drug data pricing, Part D (the new drug benefit), Medicare Advantage, and Medicare claims and payments. These databases are distinct warehouses of electronic data that reside on the mainframe. For example, the enrollment database has information on every person enrolled in Medicare since 1965 and is updated on a daily basis, including via the receipt of data files from the Social Security Administration. The claims and payment database processes millions of claims per day. With regard to the claims data files only, the mainframe processes approximately one billion claims per year. CMS was recently required to shut down the mainframe for the uploading of a database containing all Part D enrollees. As mission critical projects such as Part D move forward, additional mainframe resources will be dedicated to the project, thus slowing down other databases or systems which are on the mainframe and potentially interrupting the uploading of historical databases.

4. DESY, a data extract system, is one of the 350 systems that is run on the mainframe. DESY runs concurrently with all other CMS systems which are run on the mainframe. As an extract system to support informational, analytical, management and research needs, the DESY system is allocated far fewer resources than CMS mission critical operational

databases or systems on the mainframe. DESY supports CMS and to a much lesser extent, non-CMS needs by providing a single data extraction tool to extract data from the Medicare claims utilization and enrollment files. However, the informational requirements supported by DESY may be slowed down or delayed depending on the magnitude of CMS' day to day operational business that is running on the mainframe.

5. To give priority to a DESY project, to the detriment of the business of the Medicare program, is not an acceptable option. The mainframe is responsible for the running of claims payment systems and enrollment systems. To shut the mainframe down, or the very least slow down other applications, to run a non-critical research project could delay health benefits and services to the elderly and disabled and would certainly delay claims processing, and thus payments to providers.

6. Within DESY itself, eighty-seven percent (87%) of what DESY does is agency business, e.g., track information for fraud and abuse purposes, track program utilization and provide statistics for CMS' use. This work is critical to the integrity of the Medicare program, and the CMS is opposed to giving priority to non-government projects and has not done so in the past. The remaining capacity is used for outside projects as they can be accommodated.

7. On a weekly basis, DESY typically receives and processes between 100 to 200 agency and outside business requests which generate between 1 and 2 billion records of output each week. When a DESY request is run on the mainframe, the mainframe's limited resources which are designated for DESY use are applied to more recent years because 70% of all DESY research requests are for the most recent five years. The mainframe runs the claims data on a continuous loop. A request for a certain years data goes into a queue with other similar requests. If the mainframe is running claims data from 1998-2002, then all requests in the queue will be

processed for those years before the search can begin for claims data in 1994-1998. The limited mainframe capacity which is designated for DESY could not support the concurrent running of 15 years of claims data. In fact, if DESY attempted to search 15 years of data at the same time, it could not only bring to a halt other work, but could crash the entire mainframe. These are among the reasons why a DESY research request must be addressed in accordance with the protocol that has been developed at DMUD for handling such requests.

8. For Medicare claims there are seven separate types of claims, each which requires its own database. The seven types of claims are: home health; skilled nursing; hospital inpatient; carrier; durable medical equipment regional carrier; hospital outpatient; and clinical laboratory. Each calendar year of each of the seven claims types is another separate database. These claim databases are uploaded on the mainframe on a continuing and sequential basis. For example, at any given time the mainframe may be running four years of claims data (i.e., 7 claims types x 4 years for a total of 28 databases). Typically, since the most recent claims data has the greatest demand, that data is the most likely to be running at any given time. This is why CMS has emphasized that to the extent data for the most recent claims years can be used, that data is the most accessible.

9. Processing the data extraction request made in this case will take substantial time and resources. We can start (requests for searches tend to spike from late July through October) the search process immediately, but the nature of this request is what will slow down the process. Because we have been asked to search by J-code, each field of the electronically stored claim will have to be read sequentially. Further, we are being asked to search fifteen years of claims, with no one year, or four year increment, taking priority. Therefore, whether the mainframe is running 1998-2002, or 1994-1998 will not matter because for this request it will

eventually have to cycle through all fifteen years and search all three claims types and individual claims for that year.

10. In general, the process by which CMS could determine the names and addresses of beneficiaries who filed claims with Medicare for particular drugs involves two steps. The reason two separate steps are required is that claims data does NOT include name and address information on beneficiaries. First, the claims databases would need to be searched. The drug claims at issue would be on the outpatient, durable medical equipment, and carrier claims records within the databases. Data extraction runs would have to be performed on all three records, in four year increments for all fifteen years. In order to initiate a DESY search, a file containing all of the drug J-codes at issue and the search parameters for those J-codes would be put into the DESY queue. The mainframe determines how it will pick up and process those requests based on capacity and which years of data are currently being searched. There is no option for giving priority to any particular DESY request. The reason for this is the scheduling for the database or record runs on the mainframe is set. Also, because outside DESY requests make up only 13% of the DESY searches, batching one search by itself, and in particular the search at issue, will not speed up the search process.

11. Typically, a DESY request consists of manually asking the system for a finder file using certain criteria and search parameters. In this case, the search criteria would be the J-code, and the search parameters would be claim type and year. The finder file would consist of the numerical Medicare beneficiary identifiers, referred to as the Health Insurance Code number, or HICN. Although multiple years of claims types can be searched at once, the system can only search one type of claim, e.g., outpatient, or durable medical equipment or carrier, at a time. In order for DESY to search the billions of claims that have been processed since 1991, it must read

each claim field separately and sequentially. For example, our claims search here is for the purpose of locating certain J-codes. J-code information is buried in one of multiple fields on the claim so the system must read the entire claim to recognize that information. Once it recognizes the J-code as one of the search criteria, DESY will pull the HICN on the claim and accumulate it to the finder file.

12. DESY can only pull data from what is currently being run on the mainframe. Likely, at least half of the most recent eight years of claims data will be running on the mainframe. It takes approximately three to four weeks to search a year's worth of claims data for the most recent five years. The greater the number of years being searched at the same time, the slower the DESY search proceeds. Because of its limited capacity, DESY is unable to perform a search on 15, or even 10 years of claims simultaneously. For DESY to pull HICNs from those four to eight years of claims would take approximately three months. If CMS is required to search all 15 years of claims data, the process could take at least 6 months because of the infrequency with which the older data is run on the mainframe and the number of claims runs which would have to be performed. The time gets longer depending on how many separate runs DESY is required to perform.

13. Our search time estimates are based on experience. However, trying to guarantee that a particular search will take place in a particular time frame is impossible because of the uncertainty of the demands that could be placed on the mainframe in the future. For example, when hurricane Katrina hit the gulf states last year those states lost vital Medicaid and Medicare records. In order to locate the beneficiary populations, i.e., names and addresses, and provide important and lifesaving benefits, the CMS mainframe shut down non-essential applications such as DESY so that it could respond swiftly to the data requests for the gulf states. A further

example is that in fall 2006, the mainframe will shut down, or at the very least cease non-essential applications, in order to handle the new enrollees for the Medicare prescription drug program, or Part D.

14. Once all 15 years of claims have been searched, CMS would merge and de-duplicate the HICNs from all 45 runs (15 years x 3 claims databases) into one finder file. The merge and de-duplicate process for the 45 runs would take approximately two weeks. If CMS was required to merge and de-duplicate a larger number of runs it could take from several weeks to several months. Again, the uncertainty in the time frames is caused by the uncertainty in the demands on the mainframe. Also, the number of individual runs which would have to be merged and de-duplicated adds to the processing time. To the extent it is determined that only one name and address file is required, and not individual name and address files for each drug, the process is faster.

15. Once the merge and de-duplicate process is finished, the second step is to perform a DESY run of the finder file against CMS' enrollment and eligibility database (again, this activity takes place on the mainframe) and produce a file with the names and addresses of Medicare beneficiaries who had claims filed on their behalf for certain Part B reimbursed drugs. The enrollment and eligibility database has information on every person who was ever a Medicare beneficiary, or approximately 92 million persons. The time to run a finder file will be considerably lengthened depending on the number of HICNs on the finder file and the number of finder files. For one finder file with approximately three million HICNs, a typical run against the enrollment and eligibility database would take one to two weeks. Because we will not know the number of HICNs until after the finder file is created, the time required to run the finder file against the enrollment and eligibility database is uncertain.

16. Once all of the runs were complete, the CMS would produce from one (if no matching of drug to beneficiary was required) to 47 (if multiple J-codes for each drug were used and had to be matched to a particular beneficiary) text files with names and addresses on a CD.

17. We have been asked to estimate the cost and time it would take for CMS to run data using a few different scenarios. Below are the time and cost estimates:

a. The first scenario is the length of time and cost to find out the name and address information for each beneficiary on whose behalf a claim was filed with Medicare for at least one of the J-codes at issue (i.e., during the search at least one claim shows up with a particular J-code), **but without** matching the particular drug to the beneficiary. The estimated length of time to search 15 years of claims data is six months. The estimated length of time to run only the most current eight years of data is three months. CMS could produce the data on a rolling basis, year by year, but it would not have gone through the de-duplication process so the finder file might have multiple identical HICNs. The cost of producing the data would be \$109,050 plus the cost of the name and address information. The cost can be broken down as follows: carrier data is \$5,190 per year x 15 years = \$77,850; outpatient data is \$875 per year x 15 years = \$13,125; and DME data is \$1,205 per year times 15 years = \$18,075. Finally the cost of the name and address information is \$3,075 for each three million increment. The cost estimates are based on a formula using the amount of work the central processor unit will perform, analyst time and the cost of the media, i.e., tapes.

b. The second scenario is the length of time and cost to find out the name and address information for each beneficiary on whose behalf a claim was filed with Medicare for at least one (i.e., during the search at least one claim shows up with a particular J-code) of the

drugs at issue, **and** matching the particular drug to the beneficiary. The estimated length of time to search 15 years of claims data, using one J-code for a given drug and pulling both the HICN and the matching J-code is eight months. The estimated length of time to run only the most current eight years of data is five months. The reason for this longer estimate is that instead of 45 claim runs, CMS would now have to do 1,215 claim runs (45 times the 27 drugs at issue) resulting in 27 separate name and address files. The cost of producing the data would be \$2,944,350 plus the cost of the name and address information. The cost can be broken down as follows: carrier data is \$5,190 per year x 15 years x 27 claim runs = \$2,101,950; outpatient data is \$875 per year x 15 years x 27 claim runs = \$354,375; and DME data is \$1,205 per year x 15 years x 27 claim runs = \$488,025. Finally the cost of the name and address data is \$3,075 for each three million increment.

c. Third, I have also estimated the the length of time and cost to find out the name and address information for each beneficiary on whose behalf a claim was filed with Medicare for at least one of the drugs at issue, **and** matching the particular drug **and dosage** to the beneficiary. The estimated length of time to search 15 years of claims data, using multiple J-codes for a given drug and pulling the HICN and the matching J-code is 10 months. The estimated length of time to run only the most current eight years of data is six months. The reason for this estimate is that instead of 45 claim runs, CMS would now have to do 2,115 claim runs (45 times the 47 codes at issue) resulting in 47 separate name and address files. The cost of producing the data would be \$5,125,350 plus the cost of the name and address information. The cost can be broken down as follows: carrier data is \$5,190 per year x 15 years x 47 claim runs = \$3,658,950; outpatient data is \$875 per year x 15 years x 47 claim runs = \$616,875; and DME

data is \$1,205 per year x 15 years x 47 claim runs = \$849,525. Finally the cost of the name and address information is \$3,075 for each three million increment.

18. Assuming that the Court and the parties soon agree upon a particular search scenario (see paragraph 17 above), due to the relatively small number of data requests currently in the queue, CMS can begin to compile the information sought by the subpoena immediately.

19. On June 30, 2006, during a telephone conference call which included me, the U.S. Attorney's Office, counsel for the plaintiffs and defendants, and representatives of Complete Claims Solutions, the vendor who would issue class notice, I explained the format in which CMS would produce the requested data. Specifically, I addressed Complete Claims Solutions' questions concerning the mechanical usage of the data that CMS will collect and deliver. We agreed that CMS would deliver the data to Complete Claims Solution in a text format along with a data record layout that would facilitate the company's ability to make mailing labels for the intended mail out. The quality of the data provided by CMS should be sufficient to achieve at least 95% successful mail delivery (i.e., deliverable to a mailable address).

I declare under penalty of perjury that the foregoing is true and correct. Executed this 30th day of June, 2006 in Baltimore, Maryland.

/s/ John A. Fletcher